

Requested Patent: JP2003256455A

Title:

XML DOCUMENT STORAGE/RETRIEVAL DEVICE, XML DOCUMENT
STORAGE/RETRIEVAL METHOD USED IN IT, AND PROGRAM FOR IT ;

Abstracted Patent: JP2003256455 ;

Publication Date: 2003-09-12 ;

Inventor(s): KOIZUMI SEIICHI; TOMIZAWA NOBUYUKI; FUJITA SATORU ;

Applicant(s): NEC CORP ;

Application Number: JP20020059648 20020306 ;

Priority Number(s): ;

IPC Classification: G06F17/30; G06F12/00 ;

Equivalents: ;

ABSTRACT:

PROBLEM TO BE SOLVED: To provide an XML document storage/retrieval device capable of generating SQL inquiry according to RDB schema structure by reducing the storage cost for the XML document and the retrieval cost for the XML document. SOLUTION: XML-RDB mapping rule generation performing means 12 generates an XML-RDB mapping rule for storing XML document element not to be retrieved as it is from the XML schema description in an XML schema description language and the structural information of the XML document element to be retrieved. An XML document storage means 13 stores the XML document on the basis of the mapping rule, and an XML document retrieving means 14 generates retrieval, XML inquiry-SQL transformation rule to thereby efficiently store/retrieve the XML document.

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号
特開2003-256455
(P2003-256455A)

(43)公開日 平成15年9月12日(2003.9.12)

(51)Int.Cl. ⁷	識別記号	F I	キーワード [*] (参考)
G 0 6 F 17/30	2 4 0	C 0 6 F 17/30	2 4 0 C 5 B 0 7 D
	1 4 0		1 4 0 5 B 0 8 2
	1 8 0		1 8 0 D
12/00	5 1 3	12/00	5 1 3 A
			5 1 3 D

審査請求 未請求 請求項の数21 O L (全 16 頁)

(21)出願番号 特願2002-59648(P2002-59648)

(22)出願日 平成14年3月6日(2002.3.6)

(71)出願人 000004237
日本電気株式会社
東京都港区芝五丁目7番1号
(72)発明者 小泉 清一
東京都港区芝五丁目7番1号 日本電気株
式会社内
(73)発明者 富沢 伸行
東京都港区芝五丁目7番1号 日本電気株
式会社内
(74)代理人 100088812
弁理士 ▲柳▼川 信

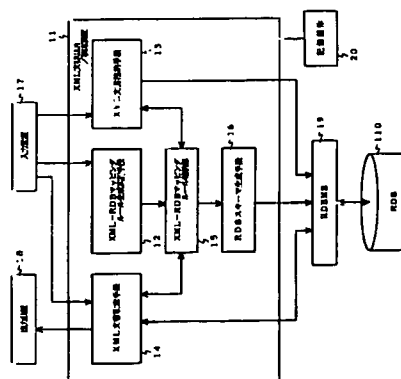
最終頁に続く

(54)【発明の名称】 XML文書格納／検索装置及びそれに用いるXML文書格納／検索方法並びにそのプログラム

(57)【要約】

【課題】 XML文書の格納コスト及びXML文書の検索コストを減少させ、RDBスキーマ構造に合わせたSQL問合せの生成を実現可能なXML文書格納／検索装置を提供する。

【解決手段】 XML-RDBマッピングルール生成実行手段12はXMLスキーマ記述言語によるXMLスキーマ記述と検索対象となるXML文書要素の構造情報とから、検索対象とならないXML文書要素をXML文書要素のまま格納するというXML-RDBマッピングルールを生成する。XML文書格納手段13はそのマッピングルールを基にXML文書の格納を行い、XML文書検索手段14は検索、XML問合せ-SQL変換ルールの生成を行うことによって、XML文書を効率的に格納／検索する。



【特許請求の範囲】

【請求項1】 XML (extensible markup language) 文書の少なくとも構造を示す前記XML文書のスキーマとRDB (relational database) の少なくとも構造を示す前記RDBのスキーマとを基に検索対象のXML文書の要素であるXML文書要素と前記RDBのスキーマとの対応付けと非検索対象のXML文書の要素であるXML文書要素をそのまま前記RDBに格納するという条件とを含むマッピングルールを生成するXML-RDBマッピングルール生成実行手段と、前記マッピングルールを利用して前記XML文書の格納を行うXML文書格納手段とを有することを特徴とするXML文書格納/検索装置。

【請求項2】 前記検索対象のXML文書要素が検索時に必要な情報であり、前記非検索対象のXML文書要素が検索後に利用する情報であることを特徴とする請求項1記載のXML文書格納/検索装置。

【請求項3】 前記XML-RDBマッピングルール生成実行手段は、前記XML文書のスキーマの構文解析を行うXMLスキーマ解析手段と、前記検索対象のXML文書要素の構造情報に対して構文解析を行う検索対象解析手段と、格納するXML文書のスキーマと前記検索対象のXML文書要素の構造情報とから前記検索対象のXML文書要素と前記非検索対象のXML文書要素とに分類する文書要素分類手段と、前記XML文書要素の分類結果から前記RDBのスキーマを構成するRDBスキーマ構成手段と、前記RDBのスキーマと前記XML文書の要素の分類結果とから前記マッピングルールを導出するXML-RDBマッピングルール導出手段とを含むことを特徴とする請求項1または請求項2記載のXML文書格納/検索装置。

【請求項4】 前記XML文書格納手段は、入力されたXML文書の構文解析を行うXML文書構文解析手段と、前記マッピングルールを基に受取った構文解析後のXML文書を前記検索対象のXML文書要素と前記非検索対象のXML文書要素とに分別しかつ前記検索対象のXML文書要素から前記RDBのスキーマに対応付けられた情報を取り出す検索対象判定手段と、前記検索対象判定手段において分別されたXML文書要素を前記RDBに格納するためのSQL (structured query language) を生成する格納命令生成手段と、前記SQLを実行する格納操作実行手段とを含むことを特徴とする請求項1から請求項3のいずれか記載のXML文書格納/検索装置。

【請求項5】 前記RDBのスキーマに対応付けられた情報がタグ名、値、属性名、属性値であることを特徴とする請求項4記載のXML文書格納/検索装置。

【請求項6】 前記マッピングルールを利用して前記XML文書の検索と再構成を行うXML文書検索手段を

含むことを特徴とする請求項1から請求項5のいずれか記載のXML文書格納/検索装置。

【請求項7】 前記XML文書検索手段は、前記XML文書の検索の検索式に対して構文解析を行う検索式解析手段と、前記マッピングルールを基に前記XML文書に対する問合せからSQL (structured query language) 問合せを生成する問合せ生成手段と、前記SQL問合せを実行する問合せ実行手段と、前記問合せ実行手段の実行結果から前記マッピングルールを基に検索結果のXML文書を構成するXML文書生成手段とを含むことを特徴とする請求項6記載のXML文書格納/検索装置。

【請求項8】 XML (extensible markup language) 文書の少なくとも構造を示す前記XML文書のスキーマとRDB (relational database) の少なくとも構造を示す前記RDBのスキーマとを基に検索対象のXML文書の要素であるXML文書要素と前記RDBのスキーマとの対応付けと非検索対象のXML文書の要素であるXML文書要素をそのまま前記RDBに格納するという条件とを含むマッピングルールを生成するステップと、前記マッピングルールを利用して前記XML文書の格納を行うステップとを有することを特徴とするXML文書格納/検索方法。

【請求項9】 前記検索対象のXML文書要素が検索時に必要な情報であり、前記非検索対象のXML文書要素が検索後に利用する情報であることを特徴とする請求項8記載のXML文書格納/検索方法。

【請求項10】 前記マッピングルールを生成するステップは、前記XML文書のスキーマの構文解析を行うステップと、前記検索対象のXML文書要素の構造情報に対して構文解析を行うステップと、格納するXML文書のスキーマと前記検索対象のXML文書要素の構造情報とから前記検索対象のXML文書要素と前記非検索対象のXML文書要素とに分類するステップと、前記XML文書要素の分類結果から前記RDBのスキーマを構成するステップと、前記RDBのスキーマと前記XML文書要素の分類結果とから前記マッピングルールを導出するステップとを含むことを特徴とする請求項8または請求項9記載のXML文書格納/検索方法。

【請求項11】 前記XML文書の格納を行うステップは、入力されたXML文書の構文解析を行うステップと、前記マッピングルールを基に受取った構文解析後のXML文書を前記検索対象のXML文書要素と前記非検索対象のXML文書要素とに分別しかつ前記検索対象のXML文書要素から前記RDBのスキーマに対応付けられた情報を取り出すステップと、その分別されたXML文書要素を前記RDBに格納するためのSQL (structured query language) を生成するステップと、前記SQLを実行するステップとを含む

むことを特徴とする請求項8から請求項10のいずれか記載のXML文書格納／検索方法。

【請求項12】 前記RDBのスキーマに対応付けられた情報がタグ名、値、属性名、属性値であることを特徴とする請求項11記載のXML文書格納／検索方法。

【請求項13】 前記マッピングルールを利用して前記XML文書の検索と再構成とを行うステップを含むことを特徴とする請求項8から請求項12のいずれか記載のXML文書格納／検索方法。

【請求項14】 前記XML文書の検索と再構成とを行うステップは、前記XML文書の検索の検索式に対して構文解析を行うステップと、前記マッピングルールを基に前記XML文書に対する問合せからSQL (structured query language) 問合せを生成するステップと、前記SQL問合せを実行するステップと、その実行結果から前記マッピングルールを基に検索結果のXML文書を構成するステップとを含むことを特徴とする請求項13記載のXML文書格納／検索方法。

【請求項15】 コンピュータに、XML (extensible markup language) 文書の少なくとも構造を示す前記XML文書のスキーマとRDB (relational database) の少なくとも構造を示す前記RDBのスキーマとを基に検索対象のXML文書の要素であるXML文書要素と前記RDBのスキーマとの対応付けと非検索対象のXML文書の要素であるXML文書要素をそのまま前記RDBに格納するという条件とを含むマッピングルールを生成するXML-RDBマッピングルール生成実行処理と、前記マッピングルールを利用して前記XML文書の格納を行うXML文書格納処理とを実行させるためのプログラム。

【請求項16】 前記検索対象のXML文書要素が検索時に必要な情報であり、前記非検索対象のXML文書要素が検索後に利用する情報であることを特徴とする請求項15記載のプログラム。

【請求項17】 前記コンピュータに、前記XML-RDBマッピングルール生成実行処理を実行させる際に、前記XML文書のスキーマの構文解析を行うXMLスキーマ解析処理と、前記検索対象のXML文書要素の構造情報に対して構文解析を行う検索対象解析処理と、格納するXML文書のスキーマ情報と前記検索対象のXML文書要素の構造情報とから前記検索対象のXML文書要素と前記非検索対象のXML文書要素とに分類する文書要素分類処理と、前記XML文書要素の分類結果から前記RDBのスキーマを構成するRDBスキーマ構成処理と、前記RDBのスキーマと前記XML文書要素の分類結果とから前記マッピングルールを導出するXML-RDBマッピングルール導出処理とを実行させることを特徴とする請求項15または請求項16記載のプログラム。

【請求項18】 前記コンピュータに、前記XML文書格納処理を実行させる際に、入力されたXML文書の構文解析を行うXML文書構文解析処理と、前記マッピングルールを基に受取った構文解析後のXML文書を前記検索対象のXML文書要素と前記非検索対象のXML文書要素とに分別しかつ前記検索対象のXML文書要素から前記RDBのスキーマに対応付けられた情報を取り出す検索対象判定処理と、この分別されたXML文書要素を前記RDBに格納するためのSQL (structured query language) を生成する格納命令生成処理と、前記SQLを実行する格納操作実行処理とを実行させることを特徴とする請求項15から請求項17のいずれか記載のプログラム。

【請求項19】 前記RDBのスキーマに対応付けられた情報がタグ名、値、属性名、属性値であることを特徴とする請求項18記載のプログラム。

【請求項20】 前記コンピュータに、前記マッピングルールを利用して前記XML文書の検索と再構成とを行うXML文書検索処理とを実行させることを特徴とする請求項15から請求項19のいずれか記載のプログラム。

【請求項21】 前記コンピュータに、前記XML文書検索処理を実行させる際に、前記XML文書の検索の検索式に対して構文解析を行う検索式解析処理と、前記マッピングルールを基に前記XML文書に対する問合せからSQL (structured query language) 問合せを生成する問合せ生成処理と、前記SQL問合せを実行する問合せ実行処理と、その実行結果から前記マッピングルールを基に検索結果のXML文書を構成するXML文書生成処理とを実行させることを特徴とする請求項20記載のプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明はXML文書格納／検索装置及びそれに用いるXML文書格納／検索方法並びにそのプログラムに関し、特にXML (extensible markup language) 文書の格納とその検索とを行う方法に関する。

【0002】

【従来の技術】 従来、XML等の構造化文書を格納／検索する場合には、RDB (relational database: リレーショナルデータベース) を利用する方法が考えられている。

【0003】 その1つとして、任意の構造を持つXML文書を格納するための手法が、アイ・イー・イー・イー・データ・エンジニアリング誌の第22巻第3号の27～34頁 (IEEE Data Engineering Bulletin, Vol. 22, No. 3, p. 27-34, Sep. 1999) や、情報処理学会論文誌: データベース、第40巻SIG3 (TOD3)

号、115～131頁、あるいは特開2001-34619号公報、または特開2001-318935号公報等に記載されている。この手法ではXML文書の構造と値とを区別してテーブルに格納することによって、任意の構造を持つXML文書の格納に対応している。

【0004】また、特定の文書構造を持つXML文書を格納する手法が存在する。入力するXML文書のXMLスキーマ構造とリレーショナルスキーマとのマッピングを行うことによって、格納するXML文書に対してリレーショナルスキーマを生成することができる。

【0005】マッピングルールを外部から入力する方法としては、特開2001-34513号公報に開示された技術があり、マッピングルールをスキーマを基に生成する手法としては、特開2000-122903号公報に開示された技術がある。

【0006】XML文書は利用用途によって、検索時に必要な情報と、検索結果として取出した後に利用する情報とに分類することができる。例えば、

```
<report_info>
<author>foo</author>
<abstract>This report describes . . . .</abstract>
<urls>
<url>http://. . . .</url>
<url>ftp://. . . .</url>
</urls>
</report_info>
```

というような報告書についての要約を記述したXML文書の格納/検索について考える。この場合、<author>要素の値としては著者の名前が記述され、<abstract>要素の値としては報告書の要約が記述されている。

【0007】そして、<urls>要素の子要素<url>には、報告書が置かれているURL (uniform resource locator) が記述されている。この時、著者名や要約は検索対象として有用であるが、URLは検索した後に報告書を取得するために必要となるデータであって、検索時の必要性は無い。

【0008】ゆえに、<author>要素と<abstract>要素とは検索時に必要な情報であり、<urls>要素は検索結果として取出した後に利用する情報であるというように分類できる。

【0009】文書構造を予め規定することによって、XML文書格納/検索装置としての汎用性は損なわれるが、特定の用途に対する利便性は向上する。そのような用途において、従来技術ではXML文書を効率的に格納/検索していない。

【0010】

【発明が解決しようとする課題】上述した従来のXML文書格納/検索方法では、文書全体を構文解析するコス

トと、構文解析後の全文書要素を格納するコストとがかかるため、XML文書の格納コスト（格納処理時間やその格納処理によるリソースの占有時間）が大きいう問題がある。この場合、格納するXML文書の特長を活かして格納コストを低減する既存技術は無い。

【0011】また、従来のXML文書格納/検索方法では、文書全体を検索対象としているため、XML文書の検索コスト（検索処理時間やその検索処理によるリソースの占有時間）が大きいう問題がある。この場合、格納するXML文書の特徴を活かして検索コストを低減する既存技術は無い。

【0012】さらに、従来のXML文書格納/検索方法では、XML文書に対するXML問合せ言語とリレーショナルデータモデルに対するSQL (structured query language) を自動的に変換する必要があるのに対して、系統的な方法が示されていないため、格納するXML文書の構造に合わせたXML問合せ言語→SQL変換が困難であるという問題がある。

【0013】そこで、本発明の目的は上記の問題点を解消し、XML文書の格納コスト及びXML文書の検索コストを減少させ、RDBスキーマ構造に合わせたSQL問合せの生成を実現することができるXML文書格納/検索装置及びそれに用いるXML文書格納/検索方法並びにそのプログラムを提供することにある。

【0014】

【課題を解決するための手段】本発明によるXML文書格納/検索装置は、XML (extensible markup language) 文書の少なくとも構造を示す前記XML文書のスキーマとRDB (relational database) の少なくとも構造を示す前記RDBのスキーマとを基に検索対象のXML文書の要素であるXML文書要素と前記RDBのスキーマとの対応付けと非検索対象のXML文書の要素であるXML文書要素をそのまま前記RDBに格納するという条件とを含むマッピングルールを生成するXML-RDBマッピングルール生成実行手段と、前記マッピングルールを利用して前記XML文書の格納を行うXML文書格納手段とを備えている。

【0015】本発明によるXML文書格納/検索方法は、XML (extensible markup language) 文書の少なくとも構造を示す前記XML文書のスキーマとRDB (relational database) の少なくとも構造を示す前記RDBのスキーマとを基に検索対象のXML文書の要素であるXML文書要素と前記RDBのスキーマとの対応付けと非検索対象のXML文書の要素であるXML文書要素をそのまま前記RDBに格納するという条件とを含むマッピングルールを生成するステップと、前記マッピングルールを利用して前記XML文書の格納を行うステップとを備えている。

【0016】本発明によるXML文書格納／検索方法のプログラムは、コンピュータに、XML (extensible markup language) 文書の少なくとも構造を示す前記XML文書のスキーマとRDB (relational database) の少なくとも構造を示す前記RDBのスキーマとを基に検索対象のXML文書の要素であるXML文書要素と前記RDBのスキーマとの対応付けと非検索対象のXML文書の要素であるXML文書要素をそのまま前記RDBに格納するという条件とを含むマッピングルールを生成するXML-RDBマッピングルール生成実行処理と、前記マッピングルールを利用して前記XML文書の格納を行うXML文書格納処理とを実行させている。

【0017】すなわち、本発明のXML文書格納／検索装置は、XML (extensible markup language) スキーマ記述言語によるXMLスキーマ記述と、検索対象のXML文書要素の構造情報とから、非検索対象のXML文書要素をXML文書要素のまま格納するというXML-RDB (relational database: リレーショナルデータベース) マッピングルールを生成している。つまり、XML-RDBマッピングルールは検索時に必要な情報である検索対象のXML文書要素とRDBのスキーマとの対応付けと、検索後に利用する情報である非検索対象のXML文書要素をそのままRDBに格納するという条件とからなる。

【0018】また、本発明のXML文書格納／検索装置では、このXML-RDBマッピングルールを基にXML文書の格納、検索、XML問合せからSQL (structured query language) の生成を行っている。

【0019】これによって、本発明では、格納されるXML文書の構造と、検索対象となるXML文書要素とが規定されているという状況において、XML文書を効率的に格納／検索する際の問題点 (上述した問題点) を解決し、XML文書の格納コスト及びXML文書の検索コストを減少させ、RDBスキーマ構造に合わせたSQL問合せの生成を実現することが可能となる。

【0020】

【発明の実施の形態】次に、本発明の実施例について図面を参照して説明する。図1は本発明の一実施例によるXML文書格納／検索装置の構成を示すブロック図である。図1において、XML文書格納／検索装置11は主にコンピュータからなり、XML-RDBマッピングルール生成実行手段12と、XML文書格納手段13と、XML文書検索手段14と、XML-RDBマッピングルール格納部15と、RDBスキーマ生成手段16と、記録媒体20とから構成されている。

【0021】また、XML文書格納／検索装置11はキーボード等の入力装置17とディスプレイ装置や印刷装

置等の出力装置18とに接続されるとともに、RDB110を管理するリレーショナルデータベースマネジメントシステム (RDBMS) 19を介して、リレーショナルテーブルを格納するRDB110に接続されている。

【0022】XML-RDBマッピングルール生成手段12は格納するXML文書のスキーマ情報と検索対象となるXML文書要素の構造情報とからXML-RDBマッピングルール (以下、マッピングルールとする) を生成して実行する。ここで、XML文書のスキーマとはXML文書要素の入れ子、繰返し等のXML文書の構造、データ型、制約等を定義するものであり、またRDBスキーマとはRDB中のテーブルの構成、データ型、制約等を定義するものである。

【0023】マッピングルールは検索時に必要な情報である検索対象のXML文書要素とRDBのスキーマとの対応付けと、検索後に利用する情報である非検索対象のXML文書要素をそのままRDBに格納するという条件とを含む。この場合、検索対象のXML文書要素とRDBのスキーマとの対応付けとは、例えばXML文書のタグ名、値、属性名、属性値と、RDBのテーブル名、属性名との対応付けを示す。

【0024】XML文書格納手段13はXML文書を格納し、XML文書検索手段14はXML文書を検索してその検索結果を出力する。XML-RDBマッピングルール格納部15はマッピングルールを格納し、RDBスキーマ生成手段16はマッピングルールからRDBスキーマを生成する。

【0025】記録媒体20にはXML文書格納／検索装置11 (コンピュータ) で実行されるプログラム、つまりXML文書格納／検索装置11の各手段を実現するためのプログラムが格納されている。

【0026】図2は図1のXML-RDBマッピングルール生成実行手段12の構成を示すブロック図である。図2において、XML-RDBマッピングルール生成実行手段12はXMLスキーマ解析手段121と、検索対象解析手段122と、文書要素分類手段123と、RDBスキーマ構成手段124と、XML-RDBマッピングルール導出手段125とから構成されている。

【0027】XML-RDBマッピングルール生成実行手段12は入力装置17から、自装置に格納するXML文書のスキーマ情報と、検索対象のXML文書要素の構造情報 (以下、分類条件とする) とを受取る。

【0028】XMLスキーマ解析手段121は入力装置17から入力されたXMLスキーマ記述 (自装置に格納するXML文書のスキーマ情報) に対して構文解析を行い、その構文が正しいと解析したXMLスキーマ記述を文書要素分類手段123に渡す。検索対象解析手段122は入力装置17から入力された分類条件に対して構文解析を行い、その構文が正しいと解析した分類条件を文書要素分類手段123に渡す。

【0029】文書要素分類手段123はXMLスキーマ解析手段121及び検索対象解析手段122で構文が正しいと解析されたXML文書のスキーマ情報及び分類条件から、検索対象の文書要素と非検索対象の文書要素とに分類し、その分類結果をRDBスキーマ構成手段124に渡す。

【0030】RDBスキーマ構成手段124は文書要素分類手段123での文書要素の分類結果からRDBスキーマを構成し、そのスキーマ情報と文書要素の分類結果とをXML-RDBマッピングルール導出手段125に渡す。XML-RDBマッピングルール導出手段125はRDBスキーマ構成手段124から渡されたRDBスキーマ情報と文書要素の分類結果とからXMLスキーマとRDBスキーマとを対応付けるマッピングルールを導出し、XML-RDBマッピングルール格納部15に格納する。

【0031】図3は図1のXML文書格納手段13及びXML文書検索手段14の構成を示すブロック図である。図3において、XML文書格納手段13はXML文書構文解析手段131と、検索対象判定手段132と、格納命令生成手段133と、格納操作実行手段134とから構成されている。また、XML文書検索手段14は検索式解析手段141と、問合せ生成手段142と、問合せ実行手段143と、XML文書生成手段144とから構成されている。

【0032】XML文書格納手段13は格納するXML文書を入力として受取る。XML文書構文解析手段131は入力されたXML文書の構文解析を行い、その構文が正しいと解析したXML文書を検索対象判定手段132に渡す。

【0033】検索対象判定手段132はXML文書構文解析手段131で構文が正しいと解析されたXML文書を、XML-RDBマッピングルール格納部15に格納されたマッピングルールを基に検索対象のXML文書要素と非検索対象のXML文書要素とに分別し、検索対象のXML文書要素からRDBスキーマに対応付けられた情報(タグ名、値、属性名、属性値)を取出して格納命令生成手段133に渡し、非検索対象のXML文書要素をそのまま格納命令生成手段133に渡す。

【0034】格納命令生成手段133は検索対象判定手段132において分別されたXML文書要素(検索対象のXML文書要素及び非検索対象のXML文書要素)をそれぞれRDB110に格納するためのSQLを生成し、格納操作実行手段134に渡す。格納操作実行手段134は格納命令生成手段133で生成されたSQLを実行し、構文解析後のXML文書(検索対象のXML文書要素及び非検索対象のXML文書要素)をRDB110に格納する。

【0035】XML文書検索手段14は入力装置17から入力され、XML文書の検索を行うための検索式を受

取る。検索式解析手段141は入力装置17から入力される検索式に対して構文解析を行い、その構文が正しいと解析した検索式を問合せ生成手段142に渡す。

【0036】問合せ生成手段142はXML-RDBマッピングルール格納部15に格納されたマッピングルールを基に、検索式解析手段141から受取った構文解析後のXML文書に対する問合せ(検索式)からSQL問合せを生成し、そのSQLを問合せ実行手段143に渡す。尚、SQL問合せを生成する場合にはXML文書に対する問合せからSQL問合せを生成するための変換ルール(検索に特化したルール)を生成し、その変換ルールを基にSQL問合せを生成することも可能であり、これに限定されない。

【0037】問合せ実行手段143は問合せ生成手段142で生成されたSQLを実行する。RDB110はSQLが実行されると、その実行結果である結果集合をXML文書生成手段144に渡す。XML文書生成手段144はXML-RDBマッピングルール格納部15に格納されたマッピングルールを基に、RDB110から得た結果集合を基に検索結果のXML文書を生成し、そのXML文書出力装置18に渡す。

【0038】図4は本発明の一実施例で用いられるXMLスキーマの構造例を示す図であり、図5は本発明の一実施例で用いられるXML文書例を示す図である。この図4に示すXMLスキーマ構造及び図5に示すXML文書を入力とした場合の本発明の一実施例の動作について説明する。

【0039】図6は本発明の一実施例によるRDBスキーマとマッピングルールとを導出する処理工程を示すフローチャートである。この図6を参照して本発明の一実施例によるRDBスキーマとマッピングルールとを導出する処理工程について説明する。この処理工程は、図6に示すように、繰返し行われる。尚、図6に示す処理はXML文書格納/検索装置11が記録媒体20のプログラムを実行することで実現される。

【0040】文書要素分類手段123はXMLスキーマ構造及びXML文書が入力装置17から入力されると、XML文書のルート要素から要素の階層毎に要素の判定を行う。まず、文書要素分類手段123はXMLスキーマ解析手段121から入力されたXMLスキーマ情報の中から未判定の文書要素を1つ選択する(図6ステップS1)。

【0041】文書要素分類手段123は検索対象解析手段122から入力された検索対象となるXML文書要素の構造情報を基に、選択された文書要素が検索対象の文書要素であるか(図6ステップS2)、検索対象でかつ繰り返しの文書要素であるか(図6ステップS3)、検索対象でなくかつ子孫要素として検索対象となる要素を含むかどうか(図6ステップS4)をそれぞれ判定する。

【0042】要素が検索対象で単一の要素の場合、RDBスキーマ構成手段124は格納するための属性を生成し、XML-RDBマッピングルール導出手段125は要素とRDBスキーマ構成手段124で生成された属性との対応関係をマッピングルールとしてXML-RDBマッピングルール格納部15に保存する(図6ステップS5)。

【0043】要素が検索対象で繰返しとなる要素の場合、RDBスキーマ構成手段124が別テーブルを用意し、格納するための属性を生成し、XML-RDBマッピングルール導出手段125は要素とRDBスキーマ構成手段124で生成された属性との対応関係をマッピングルールとして保存する(図6ステップS6)。

【0044】要素が検索対象ではないが、検索対象となる要素を含む場合には、階層を1つ下げて要素判定処理を行う(図6ステップS7)。要素が検索対象ではなく、検索対象となる要素を含まない場合には、RDBスキーマ構成手段124が格納するための属性を生成し、XML-RDBマッピングルール導出手段125は要素とRDBスキーマ構成手段124で生成された属性との対応関係をマッピングルールとして保存する(図6ステップS5)。

【0045】まだ未判定の文書要素が残っている場合には、上記の要素判定処理を繰返す(図6ステップS8、S1)。この時、テーブル情報と、どの要素をどのテーブルのどの属性に対応しているかを示すマッピングルールとが作成される。

【0046】図7は本発明の一実施例におけるマッピングルールの一例を示す図である。図7においては、上記の処理工程で作成されたマッピングルールの一例を示している。このマッピングルールは図4に示すXMLスキーマの構造例に対して要素「business Entity」の属性「business Key」と要素「name」とが分類条件として与えられた場合の例を示している。この場合、マッピングルールとしては図7(a)に示すような検索対象のマッピングルールと、図7(b)に示すような非検索対象のマッピングルールとが作成される。

【0047】図7(a)においては、XML文書の「business Entity@business Key」が「business Entity」テーブルの属性「business Key」に対応し、XML文書の「business Entity-name」が「business Entity」テーブルの属性「name」に対応していることを示している。

【0048】また、図7(b)においては、XML文書の「business Entity-description」が「business Entity」テーブルの属性「description」に対応し、XML文書の「business Entity-contacts」が「business Entity」テーブルの属性「contacts」に対応していることを示している。

【0049】図8は本発明の一実施例におけるXML文書例に対するRDBスキーマの対応関係を示す図である。これら図4と図5と図8とを参照して具体的なXMLスキーマ定義と、XML文書例に対するRDBスキーマとの対応関係について説明する。

【0050】本実施例では、上述したように、要素<name>と属性<businessKey>とが検索対象であるとする、business Entity (businessKey, name, description, contacts) (図8の71参照)というRDBスキーマが導出される。

【0051】また、本実施例では、上述したようなマッピングルール、つまり、「business Entity」の属性<business Key>は、business Entityテーブルのbusiness Keyに対応する(図8の72参照)、「business Entity」の子要素nameはbusiness Entityテーブルの属性nameに対応する(図8の73参照)、「business Entity」の子要素descriptionは、business Entityテーブルのdescriptionに対応する(図8の74参照)、「business Entity」の子要素contactsは、business Entityテーブルのcontactsに対応する(図7の75参照)というマッピングルールが導出される。

【0052】図9は本発明の一実施例によるXML文書を格納する処理工程を示すフローチャートであり、図10は本発明の一実施例による検索操作の処理工程を示すフローチャートである。これら図9及び図10を参照して本発明の一実施例によるXML文書を格納する処理工程及び検索操作の処理工程について説明する。尚、図9及び図10に示す処理はXML文書格納/検索装置11が記録媒体20のプログラムを実行することで実現される。

【0053】まず、XML文書格納手段13は格納するXML文書を入力として受取り、XML文書構文解析手段131で構文解析を行い(図9ステップS11)、検索体操作判定手段132を用いてXML-RDBマッピングルールを基にXML文書の要素を検索対象の要素とそれ以外の非検索対象の要素とに分類し(図9ステップS12)、検索対象の要素からRDBスキーマに対応付けられた情報(タグ名、値、属性名、属性値)を取出す。

【0054】続いて、XML文書格納手段13は格納命令生成手段133において、分類された要素(検索対象の要素及び非検索対象の要素)をRDB110に格納するためのSQLを生成する。この場合、格納命令生成手段

段133は検索対象の要素であれば(図9ステップS13)、検索対象の要素からRDBスキーマに対応付けられた情報(タグ名、値、属性名、属性値)を取出して格納するためのSQLを生成する(図9ステップS14)。

【0055】また、格納命令生成手段133は非検索対象の要素であれば(図9ステップS13)、非検索対象の要素をそのまま(XML文書のまま)格納するためのSQLを生成する(図9ステップS15)。格納命令生成手段133はこれらステップS14、S15によって全て要素を処理すると(図9ステップS16)、ステップS14、S15で得たSQLを合成(マージ)して格納操作実行手段34に渡す(図9ステップS17)。格納操作実行手段34においては、格納命令生成手段133で生成されたSQLを実行してXML文書の格納を行う(図9ステップS18)。格納処理後のテーブル例を図7の76に示す。

【0056】一方、XML文書検索手段14は入力装置17から入力される検索式を受取ると、検索式解析手段141を用いて構文解析を行い(図10ステップS21)、問合せ生成手段142でXML-RDBマッピングルール格納部15に保存されているマッピングルールを基に生成した、正しいと解析されたXML文書に対する問合せからSQL問合せの生成を行う(図10ステップS22)。

【0057】続いて、XML文書検索手段14では問合せ実行手段143が問合せ生成手段142で生成されたSQLを実行することで、RDB110から結果集合を受取ると(図10ステップS23)、XML-RDBマッピングルール格納部15に保存されているマッピングルールを基に、XML文書生成手段144で結果集合から結果のXML文書を生成する。

【0058】この場合、XML文書生成手段144は検索対象の文書要素であれば(図10ステップS24)、XML-RDBマッピングルール格納部15に保存されているマッピングルールを基に、元のXML文書の要素を再構成する(図10ステップS25)。

【0059】また、XML文書生成手段144は非検索対象の文書要素であれば(図10ステップS24)、結果集合からXML文書の要素を取出す(図10ステップS26)。XML文書生成手段144はこれらステップS25、S26によって全て要素を処理すると(図10ステップS27)、ステップS25、S26で得たXML文書の要素から検索結果のXML文書を生成して出力装置18に渡す(図10ステップS28)。

【0060】例えば、「名前に"A"という文字列を含むXML文書を取得せよ」という問合せは、
SELECT *
FROM business Entity
WHERE name like '%A%'

というSQLに変換される。

【0061】XML文書生成手段144は、このSQLの実行結果に対して、XML-RDBマッピングルール格納部15に保存されているマッピングルールを適用して、XML文書に成型し直すことによって、図5に示すXML文書が結果として返される。

【0062】このように、本実施例では、図8に示すように、検索対象部分のみにおいて値を抽出するため、「値の抽出コスト」と「値をテーブルに格納するコスト」とが減少するので、XML文書の格納コスト(格納処理時間やその格納処理によるリソースの占有時間)を減少させることができる。

【0063】また、本実施例では、検索対象となる要素の存在する階層毎にテーブルを用意しているため、図8に示すように、検索対象となる要素の存在する階層が、全階層より少ない場合にはテーブル数が最小となり、JOIN操作の回数が減少するので、XML文書の検索コスト(検索処理時間やその検索処理によるリソースの占有時間)を減少させることができる。

【0064】さらに、本実施例では、XML-RDBマッピングルールを基に、XML文書に対する問合せから、本発明のRDBスキーマ構造に合わせたSQL問合せの生成を実現することができる。

【0065】

【発明の効果】以上説明したように本発明は、XML文書の少なくとも構造を示すXML文書のスキーマとRDBの少なくとも構造を示すRDBのスキーマとを基に検索対象のXML文書の要素であるXML文書要素とRDBのスキーマとの対応付けと非検索対象のXML文書の要素であるXML文書要素をそのままRDBに格納するという条件とを含むマッピングルールを生成し、そのマッピングルールを利用してXML文書の格納を行うことによって、XML文書の格納コスト及びXML文書の検索コストを減少させ、RDBスキーマ構造に合わせたSQL問合せの生成を実現することができるという効果が得られる。

【図面の簡単な説明】

【図1】本発明の一実施例によるXML文書格納/検索装置の構成を示すブロック図である。

【図2】図1のXML-RDBマッピングルール生成実行手段の構成を示すブロック図である。

【図3】図1のXML文書格納手段及びXML文書検索手段の構成を示すブロック図である。

【図4】本発明の一実施例で用いられるXMLスキーマの構造例を示す図である。

【図5】本発明の一実施例で用いられるXML文書例を示す図である。

【図6】本発明の一実施例によるRDBスキーマとマッピングルールとを導出する処理工程を示すフローチャートである。

【図7】(a)、(b)は本発明の一実施例におけるマッピングルールの一例を示す図である。

【図8】本発明の一実施例におけるXML文書例に対するRDBスキーマの対応関係を示す図である。

【図9】本発明の一実施例によるXML文書を格納する処理工程を示すフローチャートである。

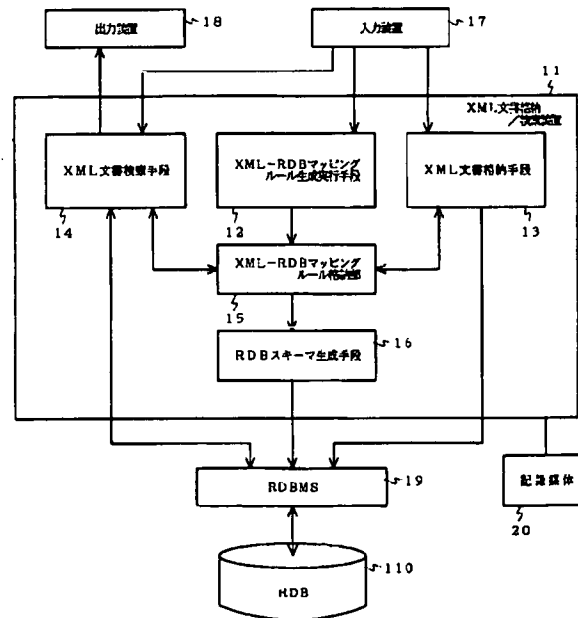
【図10】本発明の一実施例による検索操作の処理工程を示すフローチャートである。

【符号の説明】

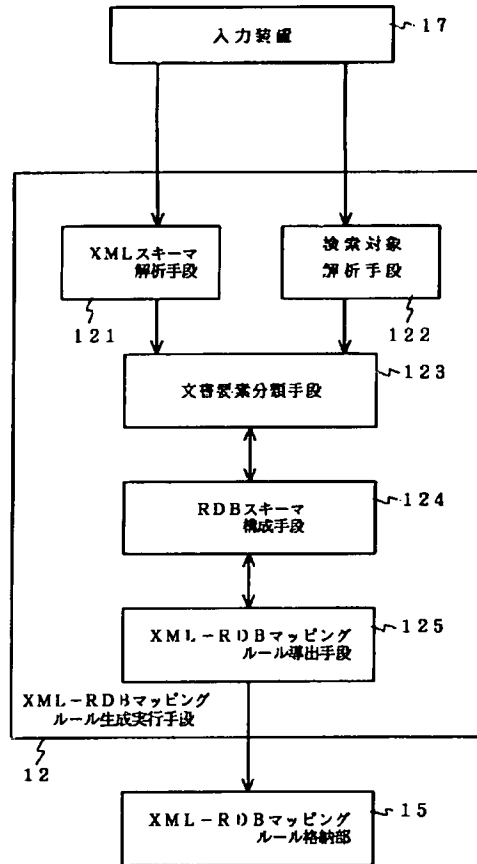
11 XML文書格納／検索装置
12 XML-RDBマッピングルール生成実行手段
13 XML文書格納手段
14 XML文書検索手段
15 XML-RDBマッピングルール格納部
16 RDBスキーマ生成手段
17 入力装置
18 出力装置
19 RDBMS
20 記録媒体

20 記録媒体
110 RDB
121 XMLスキーマ解析手段
122 検索対象解析手段
123 文書要素分類手段
124 RDBスキーマ構成手段
125 XML-RDBマッピングルール導出手段
131 XML文書構文解析手段
132 検索対象判定手段
133 格納命令生成手段
134 格納操作実行手段
141 検索式解析手段
142 問合せ生成手段
143 問合せ実行手段
144 XML文書生成手段
71 RDBスキーマ
72～75 XML-RDBマッピングルール
76 RDBテーブル例

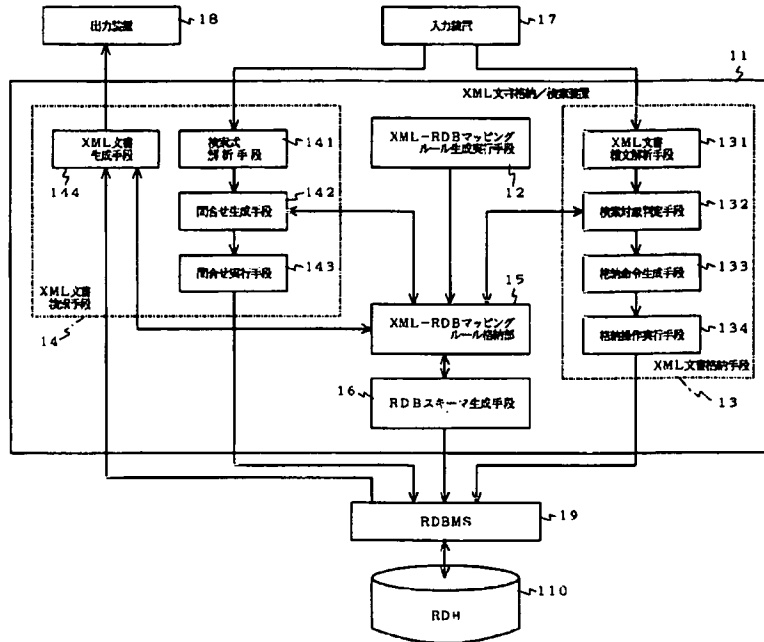
【図1】



【図2】



【図3】



【図4】

```
<business Entity business Key="XXXXXXXX-XXXX" >
  <name>A Corporations</name>
  <description xml:lang="en">A has emerged as a leading global
  Internet Solution Provider.</description>
  <contacts>
    <contact use Type="main">
      <description xml:lang="en">the address</description>
      <person Name>Corporate Headquarters</person Name>
      <phone use Type="main">(+81)X-XXXX-XXXX</phone>
      <address>7-1, Shiha 5-chome Minato-ku, Tokyo 108-8001 Japan</address>
    </contact>
  </contacts>
</business Entity>
```

【図5】

```

<schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
<element name="businessEntity"/>
<complexType name="businessEntity"><sequence>
  <element ref="name" maxOccurs="unbounded"/>
  <element ref="description" minOccurs="0" maxOccurs="unbounded"/>
  <element ref="contacts" minOccurs="0"/>
</sequence>
<attribute ref="businessKey" use="required"/>
</complexType>
<element name="contacts" type="uddi:contacts"/>
<complexType name="contacts"><sequence>
  <element ref="uddi:contact" minOccurs="0" maxOccurs="unbounded"/>
</sequence></complexType>
<element name="contact" type="uddi:contact"/>
<complexType name="contact"><sequence>
  <element ref="description" minOccurs="0" maxOccurs="unbounded"/>
  <element ref="personName"/>
  <element ref="phone" minOccurs="0" maxOccurs="unbounded"/>
  <element ref="address" minOccurs="0" maxOccurs="unbounded"/>
</sequence>
<attribute name="useType" use="optional" type="string"/>
</complexType>
</schema>

```

【図7】

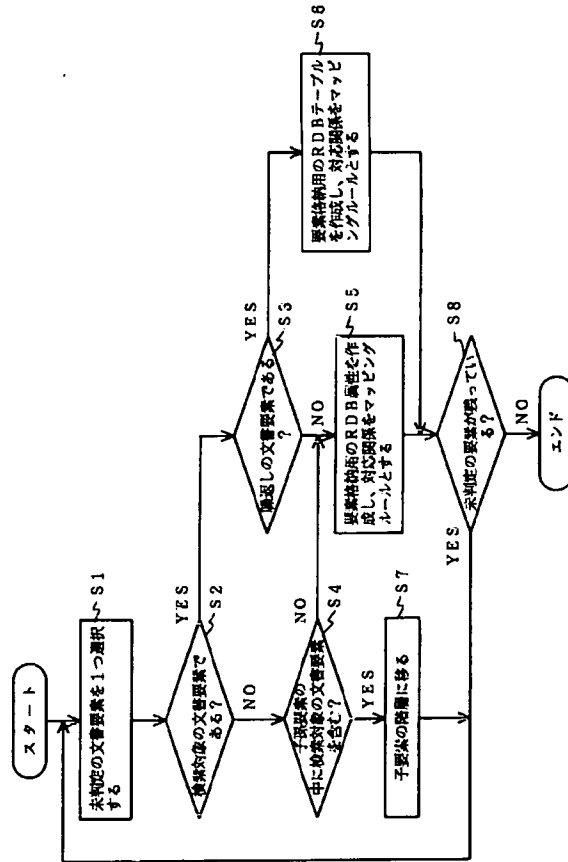
(a) 検索対象

XML	RDB
business Entity → business key	business Entity(business key)
business Entity → name	business Entity(name)

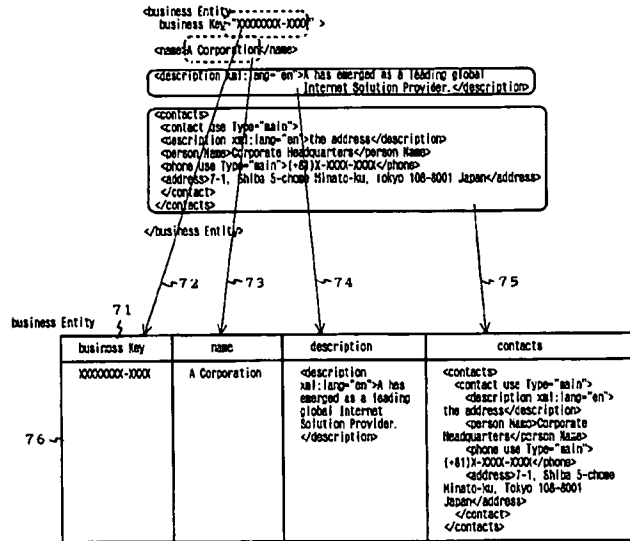
(b) 非検索対象

XML	RDB (XMLのまま)
business Entity → description	business Entity(description)
business Entity → contacts	business Entity(contacts)

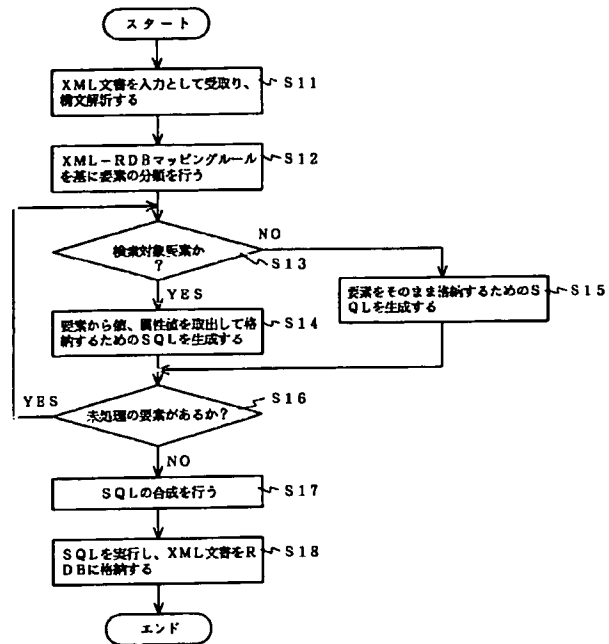
【図6】



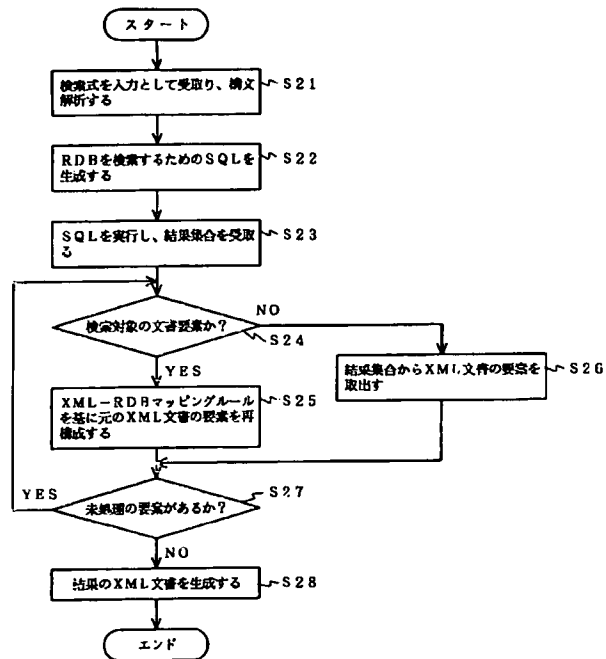
【図8】



【図9】



【図10】



フロントページの続き

(72)発明者 藤田 悟
東京都港区芝五丁目7番1号 日本電気株
式会社内

Fターム(参考) 5B075 ND02 ND34 PP26 QT06 UU06
5B082 GA07 GA08